

Understanding Information Seeking in the Patent Domain and its Impact on the Interface Design of IR Systems

Daniela Becks
University of Hildesheim
Marienburger Platz 22
D-31141 Hildesheim
+49 5121 883 -811
daniela.becks@uni-
hildesheim.de

Matthias Görtz
University of Hildesheim
Marienburger Platz 22
D-31141 Hildesheim
+49 5121 883 881
matthias.goertz@uni-
hildesheim.de

Christa Womser-Hacker
University of Hildesheim
Marienburger Platz 22
D-31141 Hildesheim
+49 5121 883 833
womser@uni-hildesheim.de

ABSTRACT

In this position paper, we highlight the need for understanding and modeling a user's context when designing the user interface of an information retrieval system using the example of the patent domain. Based on fundamental concepts of information seeking and retrieval research, we describe the different contextual factors characteristic for this domain and common task scenarios of patent retrieval. Finally, we outline first starting points for coping with these domain-specific conditions in the interface design of IR systems and discuss future research needs.

Categories and Subject Descriptors

H 1.2 [User/Machine Systems]: Human Factors; H.3.3 [Information Search and Retrieval]: Search Process. H.5.2 [Information Interfaces and Presentation]: User Interfaces – Graphical user interfaces (GUI).

General Terms

Design, Human Factors, Theory

Keywords

Information seeking; human-computer interaction; patent retrieval; context modeling

1. INTRODUCTION

While in the community of information retrieval (IR) research much effort has been spent on the invention of sophisticated system features and algorithms, the patent or intellectual property domain is still relying on Boolean systems with basic user interfaces. Even though this industry has developed highly sophisticated human strategies to seek for relevant information, one has to wonder why there has been no significant progress in terms of technical support. According to statistics of the World Intellectual Property Organization (WIPO) [14], the total number of patents in force worldwide at the end of 2006 was approximately 6.1 million. This large amount of data indicates

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HCIR '10 Workshop, August 22, 2010, New Brunswick, NJ, USA.
Copyright 2010 ACM 1-58113-000-0/00/0010...\$10.00.

that especially in this domain there should be a certain demand for innovative services to support patent users in their retrieval activities. One approach to enhancing information retrieval activities is the design of the IR system and algorithm itself. This important field of research, however, shall not be the focus of this position paper. Another approach is the design of the user interface to match the respective user need in a given context. In order to be able to develop innovative IR interfaces, it is therefore essential to understand the specificity of the domain and its impact on the interface design of IR systems.

First efforts in the field of information seeking and retrieval (IS&R) research were made. A study in the Swedish Patent and Registration Office (PRV) by Hansen & Järvelin (2005) examined the collaborative character of the patent handling process. They draw the conclusion that “future research should focus on what affects CIR (collaborative IR) processes. Possible research questions could deal with task variation, task complexity or type of task”. [7] Recently, Azzopardi et al. [1] published the results of a survey on patent users. They already started to analyze the relationship between the specialties of this kind of users, their search tasks and the functionalities of patent retrieval systems. [1]

In this position paper we would like to discuss different impact factors on information seeking, highlight the domain-specific aspects of patent retrieval, and give a first outlook on how UI design may be adapted according to these considerations.

2. THE FIELD OF IS&R RESEARCH AND ITS RELEVANCE FOR IR SYSTEMS

In the past thirty years of information science the so-called „cognitive turn“ [8] has widened the scope of traditional information retrieval (IR) research. It includes not only system-orientated perspectives of IR but also cognitive aspects of the entire process of information seeking (IS). The consideration of the actor and his context of a respective work or search task, lead to new fields of research and a variety of conceptual models on information behaviour. “Information behaviour may be defined as the more general field of investigation (as shown in Figure 1), with information-seeking behaviour being a sub-set of the field, particularly concerned with the variety of methods people employ to discover and gain access to information resources. Information searching behaviour is then defined as a sub-set of information-seeking, particularly concerned with the interactions between information user (with or without an intermediary) and computer-based information systems, of which information retrieval systems for textual data may be seen as one type.” [15]

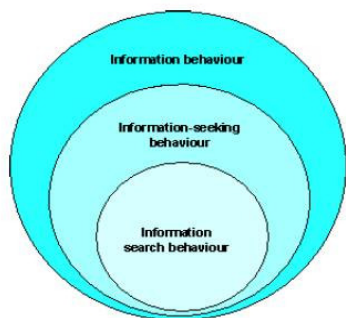


Figure 1 - Wilson's nested model of information behavior

In information seeking (IS) research, the main goal thus is, to model the user's context in order to gain a better understanding of his information needs, seeking, and use (INSU). Therein different types of models have emerged, which may be characterized as either broad or narrow, process or static, abstract or concrete, summary or analytical, general or specific models. [8] While most models of the type narrow, process, abstract, summary, and specific have been widely perceived in information science and adjacent disciplines, other more analytical considerations of the relationships of different concepts, objects, or stakeholders have found less recognition. In order to highlight domain-specific aspects of information seeking, we would like to focus on some of the most relevant findings in this area and aim to promote a deeper understanding of their impact on the design of IR systems and their interface.

In the field of IS research Leckie et al. [10] have taken such an analytical perspective on the information seeking behavior of professionals. Their basic supposition is "that the roles and related tasks undertaken by professionals in the course of daily practice prompt particular information needs, which in turn give rise to an information seeking process. However, information seeking is greatly influenced by a number of interacting variables, which can ultimately affect the outcome." [10] As their model (Fig. 2) reveals, not only the outcome is influenced by the above mentioned factors, but also the way a task and information need are perceived, and what sources of information are consulted or how aware the actor is of certain pieces of information.

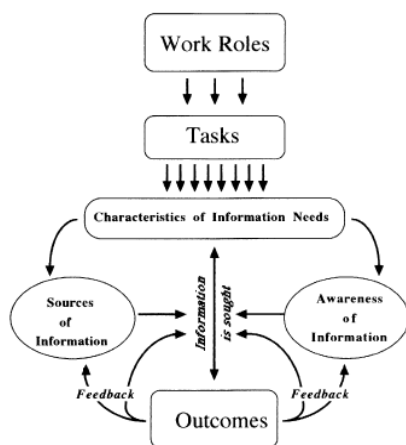


Figure 2 - The information seeking of professionals [10]

They also come to the conclusion that the way information sought is heavily influenced by "the level of complexity, the degree of importance and urgency, and whether the information need is anticipated or unexpected" [10]. The concept of task accordingly arises as central element of these findings. In a recent study by Xie [16] "the relationships between dimensions of work, search tasks, and information-seeking and -retrieving processes, in particular, the extent of planning, the application of different types of information-seeking strategies, and shifts in search-task-related goals" could be validated. Byström and Järvelin [5] also introduce the concepts of task complexity and types of information as influence factors of an actor's information seeking activities. Since task may be understood as search task as well as work task [4], the domain of work plays an important role. Since different work domains are characterized by very manifold organizational structures, areas of work, and work roles, tasks can be considered highly domain-specific.

Building upon the framework of information seeking above, we would like to highlight the importance of work role and task, and its characteristic in the patent domain in order to discuss possible conclusions for the design of user interfaces of patent retrieval systems.

3. DOMAIN-SPECIFIC ASPECTS AND TASKS OF PATENT RETRIEVAL

3.1 Description of the patent domain

The patent or intellectual property domain is characterized by a variety of domain-specific aspects, which have already been discussed by many scientists. One of these specialties is the patent document itself, because it is rather complex and contains a significant number of vague and general vocabulary [6, 9, 11]. This particularly influences the patent retrieval process, because a precise query is necessary to narrow the search and to finally find relevant documents. Furthermore, the complexity complicates the examination of a patent document, e.g. at the end of the search process. In addition to this document-specific aspect, the intellectual property domain differs from others due to the users consuming patent information. Referring to Tiwana and Horowitz [12], there is a variety of users including e.g. inventors and patent attorneys. With respect to this, Azzopardi et al. (2010) figured out that the majority of patent users are analysts or managers [1]. "In contrast to other domains such as the Web, the vast majority of practitioners of patent related retrieval are professional users." [6] This variety of user groups already implies that there might be different use cases or tasks within the patent domain. Graf and Azzopardi [6] identify the following search tasks:

- **prior art search**
Prior art search focuses on the state of the art of an invention. It is performed in order to check whether there is any existing invention similar to the one claimed in the patent to be filed (patentability search) or to invalidate a patent (invalidity search) [6]. These two types of prior art search might be seen as two separate use cases.
- **freedom-to-operate search**
A Freedom-to-Operate search aims at analyzing whether there is already a granted patent, which might be infringed by a planned product [6].

- **competitive analysis**

The third common search type is the competitive analysis. In this case, the focus is set on the patents of a competitor and it is performed to find e.g. technical information [6].

With respect to information-seeking behavior, it is recommended to consider the above mentioned use cases, because each of these search types requires a special search strategy. By now, little is known about concrete search strategies, because in the patent domain user observations are nearly impossible. Modeling information-seeking behavior on a theoretical basis can, thus, be advantageous in order to better understand patent searching and human-computer interaction.

3.2 Modeling patent-specific aspects of information seeking

According to the model developed by Leckie et al. [10] (Fig. 2), the information seeking behavior of professionals heavily depends upon the work role of the user, which corresponds to the user types in the patent domain mentioned above. The target group of patent information is already versatile, but can be even more extended. This is done by Tseng and Wu [13], who argue, that the user group can further consist of examiners, researchers and engineers. Comparing the different work roles, we might first summarize, that it can be academic/ scientific (e.g. researcher) on the one hand and purely industrial (e.g. company, experts) on the other hand. Thus, there are two general work roles that need to be considered when modeling information seeking in the patent domain. Referring to the theory of Byström [3], these are characteristic for the domain specificity of work. It should be acknowledged, that the individual role of a patent attorney is a special one, because normally he acts as an intermediary (between inventor and patent office). [11] This fact is confirmed by Azzopardi et al. [1].

Depending on the work context and the individual role of the actor, different search tasks (explained in 3.1) may be performed to fulfill the information need. For example, a competitive analysis might be especially relevant to users being involved in business decision, while an inventor should be most interested in performing a prior art search. In the patent domain, each task is “strongly shaped and driven by judicial and economic requirements” [6]. This fact further implies the influence of external factors on the work task, which in industrial contexts might, in general, be an economic interest. All in all, the patent domain is significantly task-orientated, because each search task is further dominated by the underlying goal to maximize recall and precision [6, 9], which is characteristic for this special domain.

Each search underlies an information need of the user/ actor, which, according to Leckie et al. [10], is conditioned by the defined task. In the intellectual property domain, the difficulty, with respect to the information need, is to translate it into a suitable query. As already mentioned before the language within a patent document is typically vague [6, 9, 11]. This complicates the search, because domain specific knowledge as well as intensive training is necessary to formulate an effective query and to narrow the search results returned by a retrieval system. Most of the queries are significantly complex.

Once the information need is defined, an actor has to decide which information sources to use. Within the patent domain, an

information source can be twofold, because the user has to choose between different databases on the one hand and because he has to decide which part of a patent document is most relevant to the task on the other. In case of a prior art search, for example, the claims would be the most interesting part of a patent [6]. Furthermore, to identify the state of the art of an invention, non-patent as well as patent literature should be considered during the search process [6]. These first examples indicate that the choice of information source is also influenced by the task and the information need of the actor in a given domain.

Summarizing, one can state that the patent domain provides a vivid field of application for existing analytical models in IS research and specifically contains the following influence factors on how information is sought:

- Individual work roles
- Underlying work and search tasks
- Situational and organizational factors
- Type of information need
- Task complexity
- Available information sources

4. SUMMARY AND FUTURE RESEARCH

Taking the here described observations of domain specificity in the intellectual property realm into consideration, it may be concluded that the design of the user interface of IR systems needs to account for all of the above mentioned influence factors (s. Fig. 2) in the context of the user. Next to the adaption of retrieval algorithms, the user interface presents a variety of levers to achieve this goal and make patent retrieval more effective, efficient, and more user-friendly.

Regarding the element of available **sources of information** a context-sensitive IR interface could, for example, provide integrated access to different information sources in order to assist prior art searches. Based on automatic context detection, it could also recommend suitable information sources for the different task types.

Considering the **awareness of information** interface elements for supporting query formulation or refinement could be implemented as well as the recommendation of suitable query terms. [13] Integrating query expansion methods is also suitable for narrowing the scope of a search task. With respect to this, Azzopardi et al. figured out that especially analysts would welcome such functionalities [1]. This clearly indicates that different work roles long for discriminative interface solutions.

Finally, the user interface may be designed in order to support the **outcomes** or what is referred to as “post retrieval interaction” [1] by developing innovative document views or other forms of visualizing complex document structures.

These first ideas demonstrate the relevance a deep understanding of concepts and influence factors of information seeking behavior may have on the design of user interfaces of IR systems in a specific domain of use. The cognitive viewpoint, as presented above, acts as integrating element for a variety of areas in information science [2] and has the ability to bring originally separate fields such as human-computer interaction and information retrieval together. Future research will therefore follow the line of argument of this position paper and further

analyze the impact of domain-specific factors on information seeking and the interface as well as system design of information retrieval systems. Therefore, the patent domain as well as other information-intensive domains such as the management consulting as well as academic domain will be subject to future studies of information seeking and human-computer interaction in information retrieval.

REFERENCES

- [1] Azzopardi, L., Joho, H., & Vanderbauwhede, W. (2010). Report: A Survey on Patent Users Search Behavior, Search Functionality and System Requirements. Available online: www.ir-facility.org/research/technical-reports/files/irf_tr_2010_00001.pdf
- [2] Belkin, N. J. (1990). The cognitive viewpoint in information science. *Journal of Information Science*, 16(1), 11–15.
- [3] Byström, K. (1999). *Task complexity, information types and information sources: Doctoral Dissertation*. Tampere: University of Tampere (Acta Universitatis Tamperensis 688).
- [4] Byström, K., & Hansen, P. (2005). Conceptual Framework for Tasks in Information Studies. *Journal of the American Society for Information Science and Technology*, 56(10), 1050–1061.
- [5] Byström, K., & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information Processing & Management*, 31(2), 191–213.
- [6] Graf, E., & Azzopardi, L. (2008). A methodology for building a patent test collection for prior art search. *Proceedings of the 2nd International Workshop on Evaluating Information Access (EVIA)*, December 16, 2008, Tokyo, Japan, 60–71.
- [7] Hansen, P., & Järvelin, K. (2005). Collaborative information retrieval in an information-intensive domain. *Information Processing and Management*, 41(5), 1101 - 1119.
- [8] Ingwersen, P., & Järvelin, K. (2005). *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer-11645 /Dig. Serial]: Vol. 18. Dordrecht: Springer.
- [9] Kando, N. (2000). What Shall We Evaluate? - Preliminary Discussion for the NTCIR Patent IR Challenge (PIC) Based on the Brainstorming with the Specialized Intermediaries in Patent Searching and Patent Attorneys. *ACM-SIGIR Workshop on Patent Retrieval*, July 28, 2000, Athens, Greece, 37–42.
- [10] Leckie, G. J., Pettigrew, K. E., & Sylvain, C. (1996). Modeling the Information Seeking of Professionals: A General Model Derived from Research on Engineers, Health Care Professionals, and Lawyers. *The Library Quarterly*, 66(2), 161 - 193, from <http://www.jstor.org/stable/4309109>
- [11] Schamlu, M. (1985). *Patentschriften - Patentwesen. Eine argumentationstheoretische Analyse der Textsorte Patent am Beispiel der Patentschriften zu Lehrmitteln*. München: Iudicium.
- [12] Tiwana, S., & Horowitz, E. (2009). Extracting Problem Solved Concepts from Patent Documents. *Proceedings of the 2nd ACM workshop on Patent Information Retrieval, PaIR 2009*, November 6, 2009, Hong Kong, China, 43-48.
- [13] Tseng, Y., & Wu, Y. (2008). A Study of Search Tactics for Patentability Search: A Case Study on Patent Engineers. In J. Tait (Ed.). *Proceedings of the 1st ACM workshop on Patent Information Retrieval, PaIR 2008*, October 30, 2008, Napa Valley, California, USA, 33-36.
- [14] World Intellectual Property Organization (WIPO) (2008): World Patent Report: A statistical review. Geneva: IPstats.
- [15] Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation*, 55(3), 249–270.
- [16] Xie, I. (2009). Dimensions of tasks: influences on information-seeking and retrieving process. *Journal of Documentation*, 65(3), 339–366.